A Survey on Audio Diffusion Models: Text To Speech Synthesis and Enhancement in Generative AI

Year: 2023 | *Citations: 104* | *Authors: Chenshuang Zhang, Chaoning Zhang, Sheng Zheng*

Abstract

Generative AI has demonstrated impressive performance in various fields, among which speech synthesis is an interesting direction. With the diffusion model as the most popular generative model, numerous works have attempted two active tasks: text to speech and speech enhancement. This work conducts a survey on audio diffusion model, which is complementary to existing surveys that either lack the recent progress of diffusion-based speech synthesis or highlight an overall picture of applying diffusion model in multiple fields. Specifically, this work first briefly introduces the background of audio and diffusion model. As for the text-to-speech task, we divide the methods into three categories based on the stage where diffusion model is adopted: acoustic model, vocoder and end-to-end framework. Moreover, we categorize various speech enhancement tasks by either certain signals are removed or added into the input speech. Comparisons of experimental results and discussions are also covered in this survey.